Volume 229, number 2, 273–278 FEB 05650 March 1988

Complete cDNA sequence of a *Dictyostelium* ubiquitin with a carboxy-terminal tail and identification of the protein using an anti-peptide antibody

A. Müller-Taubenberger, M. Westphal, E. Jaeger, A. Noegel and G. Gerisch

Max-Planck-Institut für Biochemie, 8033 Martinsried bei München, FRG

Received 9 December 1987; revised version received 6 January 1988

The complete sequence of a *Dictyostelium discoideum* cDNA is presented that codes for monoubiquitin extended at its C-terminus by a 52 amino acid tail. The sequence of both the ubiquitin portion and the tail is highly homologous to the one of *Saccharomyces cerevisiae* and to a partial mouse sequence. The highly basic tail sequence contains a putative metal and nucleic acid-binding motif. The gene encoding the 0.6 kb mRNA of the C-terminally extended ubiquitin is represented only once in the haploid genome. The 0.6 kb mRNA as well as its translation product, a 15 kDa protein, is expressed in exponentially growing cells and remains present for at least 5 h of development. Using antibodies against a synthetic peptide that corresponds to the C-terminal amino acid sequence, a 15 kDa protein containing the extension was also detected in yeast.

Ubiquitin; Anti-peptide antibody; (*Dictyostelium*)

1. INTRODUCTION

Ubiquitin, a highly conserved 76 amino acid protein of eukaryotic cells, is conjugated via its C-terminal glycine to target proteins [1]. In human, *Dictyostelium* and yeast cells, two types of transcripts of ubiquitin genes have been detected: mRNAs coding for polyubiquitins that are proteolytically cleaved into monomeric ubiquitin, and transcripts that code for monoubiquitin that is extended at its C-terminus by a basic peptide. In yeast two C-terminal extensions of ubiquitin have been defined on the basis of genomic DNA sequencing [2]: a 52 amino acid tail homologous to that of *Dictyostelium* [3] and to an incomplete mouse sequence [4], and a 76 amino acid tail homologous to the human fusion gene product [5]. The conserved positions of cysteine residues in the C-terminal extensions have led to the suggestion that these extensions form fingers that bind to DNA [2]. As a contribution to the elucidation of the role of ubiquitin fusion proteins we present here the complete and revised sequence of a *Dictyostelium* cDNA and provide immunological evidence for the presence of a 15 kDa protein in *D. discoideum* and *Saccharomyces cerevisiae* that carries the C-terminus of the fusion protein.

Correspondence address: A. Müller-Taubenberger, Max-Planck-Institut für Biochemie, 8033 Martinsried bei München, FRG

2. MATERIALS AND METHODS

2.1. *cDNA cloning and sequencing*

To get full-length clones, a λgt11 cDNA library [6] kindly provided by Drs R. Kessin and M.-L. Lacombe, Columbia University, was screened with a tail-specific cDNA probe that was obtained by subcloning fragment UB1C [3] in pUC19. Recombinant phages were grown on *E. coli* strain Y1088 [7] for isolation of DNA. The inserts were subcloned into vectors M13mp18/19 or pUC19 and sequenced according to Sanger et al. [8] using uni-primer or 18-mer oligonucleotide primers as indicated by blocks in fig.2. Oligonucleotides for sequencing and Northern blotting were synthesized by the phosphoamidite method [9] on an Applied Biosystems DNA synthesizer and purified by HPLC.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

2.2. *DNA and RNA isolation and hybridization*

D. discoideum strain AX2-214 was grown axenically at 21°C and starved as described [10]. For Southern blotting, DNA was isolated from purified nuclei of *D. discoideum*, digested with restriction enzymes (Boehringer Mannheim), separated on 1% agarose gels in Tris-phosphate buffer, pH 7.8 [11], transferred to nitrocellulose filters (BA85, Schleicher and Schuell), and probed with nick-translated cDNA probes in 50% formamide and 2 × SSC at 37°C [3]. Highly specific probes were obtained by subcloning fragments UB1U and UB1C [3] in pUC19.

For Northern blotting, poly(A) RNA was enriched from total cellular RNA on oligo(dT) cellulose (BRL), separated on 1.2% agarose gels in the presence of 6% formaldehyde [11], and either probed with ^{32}P-endlabeled 18-mer oligonucleotides or with nick-translated fragments UB1U and UB1C.

For primer extension according to T. Dingermann (personal communication), poly(A)-enriched RNA from growth-phase cells was hybridized to 5′-end labeled oligo(C) (fig.2), incubated with reverse transcriptase and nucleotide triphosphates, and analyzed on a sequence gel.

2.3. *Anti-peptide antibodies and immunoblotting*

The C-terminal amino acids 115–128 were linked through a glycyl spacer to an N-terminal lysine residue derivatized by two palmitic acid moieties [12]. The peptide was synthesized on a *p*-alkoxybenzylalcohol resin. The N-terminal lysine was coupled as the N_α,N_ϵ-bis-Fmoc derivative and the two fatty acid moieties were introduced by coupling of palmitic acid with dicyclohexylcarbodiimide. The dipalmitoyl hexadecapeptide was obtained by cleavage from the resin using 50% trifluoroacetic acid in methylene chloride. The sequence was verified by mass spectroscopy.

250 μg of the dipalmitoyl peptide in 1 ml emulsion of complete Freund's adjuvant were injected subcutaneously into a rabbit, followed by four intracutaneous boosts of 250–500 μg of the peptide with incomplete Freund's adjuvant. Antibodies

```
cDU17          TTTTTTTCCGTTAACCTCCTTAGAAAACAGATCAAG ATG CAA ATC TTC GTT AAA ACC TTA ACT   63
Dicty.                                              Met Gln Ile Phe Val Lys Thr Leu Thr    9
yeast                                               -   -   -   -   -   -   -   -   -     9
mouse                                               -   -   -   -   -   -   -   -   -     9

cDU17    GGT AAA ACC ATC ACT CTC GAA GTT GAA GGT TCA GAC AAC ATC GAA AAC GTC AAA GCC AAG   123
Dicty.   Gly Lys Thr Ile Thr Leu Glu Val Glu Gly Ser Asp Asn Ile Glu Asn Val Lys Ala Lys   29
yeast     -   -   -   -   -   -   -   -   -   Ser -   -   Thr -   Asp -   -   -   Ser -    29
mouse     -   -   -   -   -   -   -   -   -   Pro -   -   Thr -   Glu -   -   -   Ala -    29

cDU17    ATT CAA GAC AAA GAA GGT ATC CCA CCA GAT CAA CAA CGT CTT ATC TTT GCC GGT AAA CAA   183
Dicty.   Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln   49
yeast     -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -    49
mouse     -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -    49

cDU17    TTA GAA GAT GGT CGT ACT TTA TCC GAT TAT AAC ATC CAA AAG GAA TCA ACC CTC CAC TTA   243
Dicty.   Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu   69
yeast     -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -    69
mouse     -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -    69

cDU17    GTT TTA AGA TTA AGA GGT GGT ATT GAA CCA TCC CTC GTC ATT CTT GCC CGT AAA TAC AAA   303
Dicty.   Val Leu Arg Leu Arg Gly Gly Ile Glu Pro Ser Leu Val Ile Leu Ala Arg Lys Tyr Lys   89
yeast     -   -   -   -   -   -   -  (-)2 -   -   -   -   Lys Ala -   -   Ser -   -   Asn  90
mouse     -   -   -   -   -   -   -  (-)2 -   -   -   -   Arg Gln -   -   Gln -   -   Asn  90

cDU17    TGT GAC AAA ATG ATT TGC AGA AAA TGT TAT GCT CGT TTA CAC CCA CGT GCT GTT AAC TGT   363
Dicty.   Cys Asp Lys Met Ile Cys Arg Lys Cys Tyr Ala Arg Leu His Pro Arg Ala Val Asn Cys   109
yeast     -   -   -  Ser Val -   -   -   -   -   -   -   -   Pro -   -   -   Thr -   -    110
mouse     -   -   -  Met  *                                                               94

cDU17    CGT AAG AAG AAA TGT GGT CAC TCT AAC AAT TTA AGA CCA AAG AAG AAG TTA CTT AAA TAA   423
Dicty.   Arg Lys Lys Lys Cys Gly His Ser Asn Asn Leu Arg Pro Lys Lys Lys Leu Leu Lys Stop  128
yeast     -   -  Arg -   -   -   -  Thr -  Gln -   -   -   -   -   -   -       Lys Stop  128

cDU17    ATTCTTAAAGAAATTACTAAAAAAAGTATACTATTCCTCAAAATAATAAAAAATGTCTTTTATTTTTTTATTTAATTTC  502

cDU17    TTTAAAAAAAAAAAAAAAAAAAAAAAAAAA                                                  532
```

Fig.1. Sequence of cDNA clone cDU17 of *D. discoideum*, and comparison of the derived amino acid sequence of the corresponding DUB1 gene (Dicty.) with that encoded by the UBI1/UBI2 gene of *Saccharomyces cerevisiae* (yeast) [2] and with the mouse arf2 sequence (mouse) [4]. This latter sequence is incomplete and ends at position 95 (*). Identical positions are depicted by bars. The isoleucine in position 77 of the yeast and mouse sequence is duplicated as indicated by $(-)_2$. The number of amino acid residues in the extension of *Dictyostelium* and yeast is nevertheless the same because of an extra leucine before the terminal lysine in the *Dictyostelium* sequence. At the C-terminal end of the *D. discoideum* protein is the sequence of a synthetic peptide (underlined) against which antibodies were obtained.

were affinity purified by absorption to nitrocellulose-bound peptide and elution at pH 2.5. For immunoblotting, total cellular proteins of *D. discoideum* or exponentially growing *Saccharomyces cerevisiae* BWG1-7A were separated by SDS-polyacrylamide electrophoresis in 14% gels. Yeast homogenates prepared in a French press were kindly provided by M. Marget. Blots were incubated with the affinity purified antibodies and indirectly labeled with ^{125}I-goat anti-rabbit IgG.

3. RESULTS

3.1. *Sequence of a* D. discoideum *cDNA comprising ubiquitin plus 3'-terminal extension*

The complete coding region and flanking sequences of the cDNA clone cDU17 are shown in fig.1. The coding region encompasses 384 nucleotides that code for the 76 amino acids of ubiquitin and the 52 amino acids of its C-terminal extension. By primer extension analysis three transcription start sites upstream of the initiator AUG were identified. The 3'-nontranslated region shows two putative polyadenylation signals [3]. In order to exclude artificial rearrangements in the cDNA of clone cDU17, two other cDNA clones were analysed. The 3'-untranslated portions were shorter in these clones than in cDU17 but the sequence of the coding region was the same.

The base in position 373 was difficult to identify. In three independent sequencing runs from 5' to 3', G was read in this position and we have published a sequence assuming that this was correct [3]. This G was not found when other cDNA clones and other sequencing strategies as outlined in fig.2 were used. Eliminating the G resulted in the amino acid sequence of fig.1, which is 11 residues longer than the previously published one. In order to confirm that sequence and to demonstrate that the transcripts encoding the ubiquitin fusion protein are in fact translated, a rabbit antiserum was raised against a synthetic peptide that contains only the 14 amino acids of the C-terminus in the reading frame shown in fig.1, in order to probe for the protein.

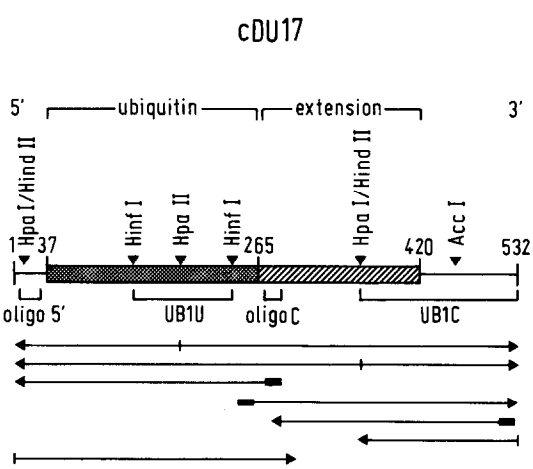

Fig.2. Map of restriction sites in the cDU17 sequence and location of subcloned fragments (UB1U and UB1C) or synthetic oligonucleotides. Oligo5' comprises nucleotides 6–30 and oligo(C) nucleotides 268–285. Blocks represent 18-mer oligonucleotides used as primers and arrows indicate directions of sequencing.

3.2. *Expression of ubiquitin fusion protein in* Dictyostelium *and yeast*

Antibodies against the C-terminal peptide of the

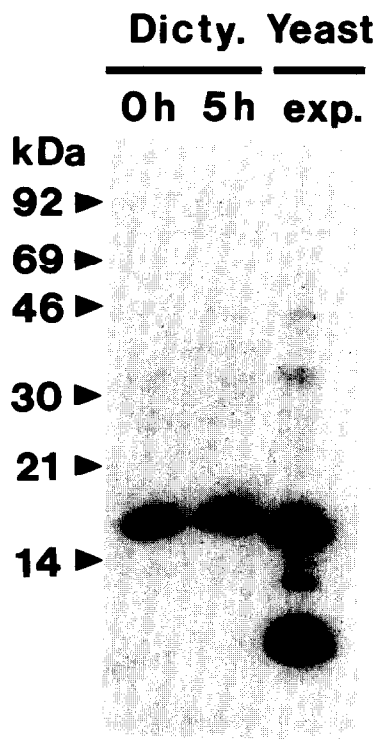

Fig.3. Proteins of *D. discoideum* (Dicty.) and of *Saccharomyces cerevisiae* (yeast) recognized by antibodies against the C-terminal peptide of the underlined *D. discoideum* sequence in fig.1. *D. discoideum* cells were harvested at exponential growth (0 h) or after 5 h of starvation, yeast cells at exponential growth (exp.). Positions of Rainbow molecular mass markers (Amersham) are indicated.

predicted *D. discoideum* sequence were affinity purified and applied to immunoblotting of total cellular proteins from *D. discoideum* and *Saccharomyces cerevisiae* (fig.3). In *D. discoideum* one protein with an apparent molecular mass of about 16 kDa was strongly labeled. In yeast two major proteins, of 15 kDa and of about half that size were recognized by the same antibodies. Since the molecular mass calculated for the entire protein from the cDNA-derived sequence is 14.7 kDa, these results suggest in both organisms the presence of proteins that consist of an ubiquitin moiety linked to the C-terminal extension.

3.3. *The 3'-extension is unique to the monoubiquitin transcript*

Northern blots were probed with recloned cDNA fragments and with oligonucleotides as indicated in fig.2. The ubiquitin-specific probe (UB1U) recognized the seven transcripts of *D. discoideum* ranging from 0.6 to 1.9 kb that have been shown to contain ubiquitin sequences [3]. In contrast, the probe specific for the 5'-untranslated region of the cDU17 sequence (oligo5'), the probe specific for the 3'-extension [oligo(C)], and the probe recognizing the extension plus the 3'-untranslated region (UB1C) hybridized only to the 0.6 kb transcript (fig.4). These results show that none of the larger transcripts, which code for polyubiquitins [13], carries the 3'-extension and they also show that no transcript encoding proteins other than ubiquitin contain a sequence homologous to the extension.

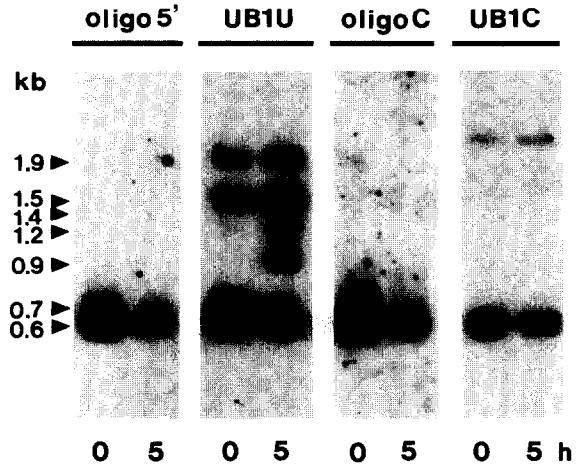

Fig.4. Northern blot analysis of transcripts recognized by various probes complementary to the cDU17 sequence. The probes range from the 5'-untranslated region up to the 3'-end as indicated in fig.2. Only the UB1U fragment that recognizes part of the ubiquitin coding region hybridized to more than a single band. The 2.0 kb band recognized by the UB1C probe is ribosomal RNA that was not completely removed from the poly(A)-enriched RNA preparation. Sizes of transcripts are taken from [14].

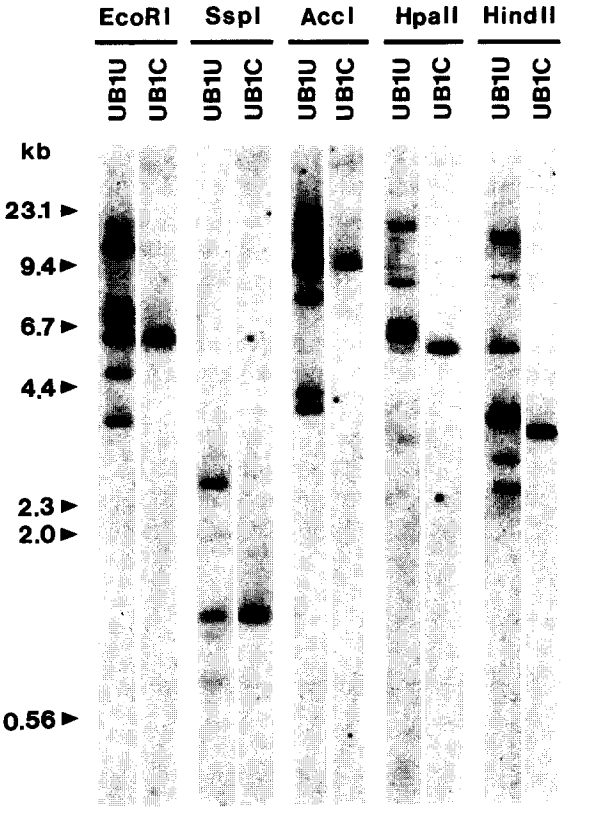

Fig.5. Southern blot analysis of genes coding for ubiquitin and for the C-terminally extended ubiquitin. Two of the restriction enzymes used, *Eco*RI and *Ssp*I, cut only outside of the cDU17 sequence, *Hpa*II within the ubiquitin coding region, *Acc*I within the 3'-untranslated region, and *Hin*dII within the region coding for ubiquitin and the C-terminal extension (fig.2). Of the two probes used, UB1U recognizes an ubiquitin encoding fragment with the *Hpa*II site in the middle, and UB1C the region downstream of the *Hin*dII site. In the blot shown the band recognized by both probes in the *Hpa*II digest was only weakly labeled by the UB1U probe. Labeling of this band seems to depend critically on hybridization stringency; in two other experiments the band was about as strongly labeled as either of the two slightly larger fragments. Positions of *Hin*dIII fragments of λDNA are indicated.

3.4. *Genomic monoubiquitin and 3'-extension sequences are linked to each other in a single-copy gene*

Southern blots of genomic DNA cut with five different restriction enzymes were hybridized with the recloned fragments, one of them recognizing the ubiquitin sequence (UB1U), the other the extension plus the 3'-untranslated portion (UB1C) (fig.5). The results obtained indicate that the transcript corresponding to the cDU17 clone is encoded by a single continuous gene, DUB1. The UB1U probe recognized up to 8 bands, in accord with the multiplicity of polyubiquitin genes [3,13]. The UB1C probe recognized only a single band in each of the restriction digests, the smallest one with a size of about 1.1 kb which is less than twice the size of the DUB1 gene (fig.1). In each digest the band recognized by the UB1C probe coincided with one of the bands recognized by the UB1U probe. An exception was the *Hin*dII digest where no coinciding bands were found. This was expected since *Hin*dII cleaves between the regions of the cDU17 sequence that are recognized by the two probes (fig.2).

4. DISCUSSION

The complete cDNA-derived sequence of a *D. discoideum* monoubiquitin with a C-terminal extension shows that not only the ubiquitin but also the extension is highly conserved in organisms as different as yeast, *Dictyostelium* and mouse [2–4]. Compared to the yeast UB11/UB12 amino acid sequence [2] there are four amino acid exchanges within the ubiquitin part of the *D. discoideum* sequence, and 11 within the extension (fig.1). Four of the exchanges in the extension are conservative amino acid substitutions. Compared to the incomplete mouse sequence [4], 2 amino acid exchanges are found within the ubiquitin part and 4 within the 18 identified amino acid residues of the extension.

Putative functions of ubiquitin fusion proteins have been discussed by Özkaynak et al. [2]. Of particular interest is the possibility that the C-terminal extension of the ubiquitin forms a DNA-binding 'finger' involving four of the five conserved cysteine residues. Detection of a protein with the full size of the fusion protein in *D. discoideum* as well as *Saccharomyces cerevisiae* cells shows that the translation product is not immediately cleaved. Thus the lifetime of ubiquitin with the C-terminal extension appears to be much longer than the persistence of the polyubiquitin precursors of monoubiquitin. In addition, a smaller protein carrying the C-terminus of the extension has been detected in yeast. It remains to be investigated whether this protein represents the extension from which the ubiquitin moiety is cleaved off, and whether cleavage is important for the protein to fulfil its function.

The 0.6 kb transcript is maximally expressed in growing cells. Its amount slightly declines during early development (fig.4), while the amount of protein recognized by the antibodies remains essentially unchanged (fig.3). Heat-shock, CD^{2+} and cycloheximide treatment, which cause a strong accumulation of polyubiquitin transcripts, leave the amount of the 0.6 kb transcripts unaffected [14]. These results suggest that the function of the 15 kDa protein that is encoded by these transcripts differs from other ubiquitin activities. Localization of the protein using the anti-peptide antibodies against the C-terminal tail may help to elucidate the function.

Acknowledgements: We thank J. Wippler for the synthesis of oligonucleotides, Drs S. Modrow and M. Schleicher for advice regarding anti-peptide antibody production and affinity purification. We are grateful to Dr M.J. Schlesinger, St. Louis, for his suggestions concerning the tail sequence. A.M.-T. acknowledges a Kekulé-Stipendium of the Fonds der Chemischen Industrie.

REFERENCES

[1] Finley, D. and Varshavsky, A. (1985) Trends Biochem. Sci. 10, 343–347.
[2] Özkaynak, E., Finley, D., Solomon, M.J. and Varshavsky, A. (1987) EMBO J. 6, 1429–1439.
[3] Westphal, M., Müller-Taubenberger, A., Noegel, A. and Gerisch, G. (1986) FEBS Lett. 209, 92–96.
[4] St. John, T., Gallatin, W.M., Siegelman, M., Smith, H.T., Fried, V.A. and Weissman, I.L. (1986) Science 231, 845–850.
[5] Lund, P.K., Moats-Staats, B.M., Simmons, J.G., Hoyt, E., D'Ercole, A.J., Martin, F. and Van Wyk, J.J. (1985) J. Biol. Chem. 260, 7609–7613.
[6] Lacombe, M.-L., Podgorski, G.J., Franke, J. and Kessin, R.H. (1986) J. Biol. Chem. 261, 16811–16817.

[7] Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80, 1194–1198.
[8] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[9] Sinha, N.D., Biernat, J., McManus, J. and Köster, H. (1984) Nucleic Acids Res. 12, 4359–4557.
[10] Malchow, D., Nägele, B., Schwarz, H. and Gerisch, G. (1972) Eur. J. Biochem. 28, 136–142.
[11] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY.
[12] Hopp, T.P. (1984) Mol. Immunol. 21, 13–16.
[13] Giorda, R. and Ennis, H.L. (1987) Mol. Cell. Biol. 6, 2097–2103.
[14] Müller-Taubenberger, A., Hagmann, J., Noegel, A. and Gerisch, G. (1987) J. Cell Sci., in press.